

Conservation and Expression of SNF2 Proteins in *Chondrus crispus*

by

Ethan Caudell

A Senior Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Ethan Caudell

Greenville, NC

May 2014

Approved by:

Faculty Mentor (signature required):



/ Dr. John Stiller

Conservation and Expression of SNF2 Proteins in *Chondrus crispus*

By Ethan Caudell

A thesis submitted to the Faculty of East Carolina University in partial fulfillment of the requirements for the degree of Honors Bachelor of Science and the Honors College in Biological Sciences with Distinction.

Advisor: John Stiller, Ph. D East Carolina University Department of Biology

Spring 2014

Conservation and Expression of SNF2 Proteins in *Chondrus crispus*

Ethan Caudell Department of Biology, East Carolina University, Greenville NC 27858

ABSTRACT-The SNF2 families of ATP-dependent helicases are the central catalytic subunits for chromatin remodeling complexes. The most widely studied is the multisubunit SWI/SNF complex, which has been found in plants, yeast, and model organisms. The SWI/SNF core subunit is a member of a smaller subgroup (SNF2 subfamily) within the larger SNF2 super-family. In a recent study of two red algae, members of most SNF2 helicase sub-families were found to be present, although it appeared unlikely that the multi-subunit SWI/SNF remodeling complexes were formed based on available gene expression data. The core SWI/SNF complex is made up of four major proteins; Swi2/Snf2, Swi3, Swp73, and Snf5. The complete genome of another red algae, *Chondrus crispus*, recently was sequenced and examined to see whether the major SWI/SNF components were conserved. Using known protein sequences obtained from model organisms, SWI/SNF complex protein sequences were searched using the BLAST tool found on the NCBI website. Paralogs from the broader SNF2 ATPase super-family were determined in *Chondrus* and carefully identified to subfamily using maximum-likelihood phylogenetic analyses. The combined bioinformatics analyses revealed that *Chondrus* contains sequences from each major group of the SNF2 protein complexes. Expression data revealed that three of the four core components of the SWI/SNF complex are expressed. Even with the lack of expression from the swi73 gene the SWI/SNF and INO80 complex were found to have

correlating expression data in 15 out of 18 replicates. This data suggest that there could be comparable chromatin remodeling complexes in red algae however, they may not include the same core proteins.

Acknowledgements

I would sincerely like to thank the people that helped to make all this possible: Dr. John Stiller, Dr. Susan McRae, Mr. Kevin Baxter, Jonas Collén, My Parents, Dohn and Luanne Caudell, my brother, Will Caudell, my friends, The Department of Biology, and East Carolina University. I would also like to thank my fellow students in my honors thesis writing class for their support, suggestions, and help along the way.

Table of Contents

	Page(s)
Introduction	8-12 Red
Algae.....	8 <i>Chondrus</i>
<i>crispus</i>	9
SNF2.....	10-12
Methods.....	13-15
Results.....	19
Discussion.....	20
Reference Page.....	21-22

List of Tables

Page(s)

Table 1 (Expression Data).....	18
--------------------------------	----

List of Figures

Page(s)

Figure 1 (image of *C. crispus*).....9

Figure 2 (Phylogenetic Tree).....16

Figure 3 (Expression data conditions).....17

Introduction

Red Algae

The Rhodophyta also known as red algae, is composed of multicellular eukaryotic algae that are typically found in temperate and colder marine environments although some are found in fresh water. Red algae lack flagella and centrioles and are believed to be the sister taxa to green algae and plants (Stiller et al. 2012). More than 6000 species have been identified. Rhodophyta are primary producers and play many essential economic roles. Red Algae are used to make agar, the laboratory used culture media, and some is grown for human consumption. Red Algae get their red color from a pigment known as phycoerythrin, which reflects red light and absorbs blue light (Introduction to Rhodophyta, 1999). Because blue light penetrates water to a greater depth than light of longer wavelengths, these pigments allow red algae to photosynthesize and live at somewhat greater depths than most other "algae". Some rhodophytes have very little phycoerythrin, and may appear green or bluish from the chlorophyll and other pigments present in them (ucmp berkely). They provide a home to other marine organisms; have multiple medical and food uses, and help to maintain coral reefs.

Rhodophytes have naturally occurring antiviral and immune enhancing agents which help strengthen the immune system and fight off illness. They are also known to reduce the level of bad cholesterol. This is one of the main reasons red algae have been used as food in China for hundreds of years. Red Algae are also a good source for minerals, carbohydrates, protein, enzymes and fiber (Introduction to Rhodophyta, 1999). Little is known about the genetic and biochemical mechanisms that govern the development in the

Rhodophyta, including why they failed to develop complex parenchyma and why tissue differentiation did not take place (Stiller et al. 2012).

Chondrus crispus

The specific species of Rhodophyta that I investigated is *Chondrus crispus*. *Chondrus crispus* commonly known as Irish moss, is of great economic use and importance (Pizzola and Rayment, 2008). Agar is a product of red seaweeds, prepared from the mucilaginous components of the cell walls of certain species, including *Chondrus*. Agar has multiple uses in the production of pharmaceuticals, cosmetics, and culture media. *Chondrus* is found in abundance along the rocky parts and in tidal pools of the Atlantic coast of North America and Europe. It is a relatively small red algae getting to only about twenty centimeters in length and typical life span is two to three years. *Chondrus* branches off into four or five fronds like a fan. The complete genome has been sequenced from *C. chondrus*, as well as a large amount of transcriptional data. This aids

 algaeBASE



with the investigation of the genes that are present and how they are expressed. (Collén et al. 2013).

Figure 1. “*Chondrus crispus*”, by Michael Guiry. Retrieved from Algaebase website,

http://www.algaebase.org/search/images/detail/?img_id=Ycb4feb0bc12bdcd1.

SNF2

The SNF2 family is a large group of helicase-like proteins. The SNF2 families of ATP-dependent helicases are the central catalytic subunits for chromatin remodeling complexes. Many of the members are core motor polypeptides within ATP-dependent chromatin remodeling complexes. SNF2 proteins typically are found in eukaryotes, but also can be found in bacteria and Archaea (Eisen, 1995). SNF2 family members fall into subfamilies typically separated based on their functions. Most of the SNF2 family proteins are DNA-dependent ATPases; however, it is still uncertain how these proteins work as enzymes and mechanical machines (Flaus et al. 2006). SNF2 ATPase's aid in nuclear functions, which includes the regulation of RNA polymerase II during transcription, elongation, termination, DNA methylation and repair, chromosome condensation, homologous recombination and RNA interference pathways (Eisen, 1995).

The SNF2 protein family includes proteins that have a variety of roles in many cellular processes. SNF2 proteins originally were identified genetically as a result of the role they played in sugar metabolism (Sucrose Non Fermentation, or SNF) (Flaus, 2006). Many of the well-studied members in the family are core motor polypeptides within ATP-dependent chromatin remodeling complexes. Some of the processes for which these proteins are responsible include transcriptional regulation, recombination and various types of DNA repair. Recent studies have shown that these proteins appear to facilitate activation by gene-specific regulatory proteins. It is also known that the SNF2s are nucleic acid-dependent ATPase's (Flaus, 2006). Mutations in the nucleoside triphosphate (NTP)-binding motif and other conserved motifs impair SNF2 function. The fact that SNF2 proteins often can be substituted by other subfamily members without impairing

their function indicates that the helicase-related domains are functionally interchangeable (Eisen, 1995).

In addition to sequence similarity with other family members, all proteins in the SNF2 family contain sequence motifs similar to those found in many DNA and RNA helicase protein families (Eisen, 1995). Members of the family have been linked to a variety of diseases and some are essential for survival. Proteins with helicase motifs have been divided into superfamilies based upon the patterns of their amino acid sequence found within the motifs. Some of the biological functions include gene-specific transcriptional activation (SNF2 subfamily), transcriptional repression (MOT1), destabilization of reconstituted nucleosomes (SNF2 and SNF2L subfamilies), nucleotide excision repair of non-transcribed regions of the genome (RAD16), recombination repair (RAD54 subfamily), and chromosome segregation (Iodestar) (Kadonaga and Pazin, 1997). Based on patterns in conserved motifs, the sequence is then assigned to a specific helicase subfamily.

Little is known about the genetic and biochemical mechanisms that govern cellular development in the Rhodophyta, including why they failed to develop complex parenchyma and why tissue differentiation does not take place (Stiller et al. 2012). During development, variation in chromatin structure and formation can be controlled by multi-subunit chromatin remodeling complexes. SWI/SNF and INO80 complexes have been found to play a role in the control of transcriptional regulation in animals, yeasts, and some plants. The SWI/SNF core subunit is a member of a smaller subgroup (SNF2 subfamily) within the larger SNF2 super-family. The SWI/SNF complex was the first complex to be discovered and is the most widely studied (Winston and Carlson 1992). Homologs from the core SWI/SNF subunit can be found in all complete plant genomes (Jerzmanowski 2007) and, through other experimental research, multiple homologs have been

shown to carry out similar functions as those found in yeast and animals (Archacki et al. 2009). The core SWI/SNF complex consists of four major proteins: Swi2/Snf2, Swi3, Snf5, and Swp73 (SWI is short for SWITCH). In a recent study of another red algae *Porphyra*, Swi3 and Snf5 were not found to be present and it was thought that that SWI/SNF complex may not be conserved across red algae (Stiller et al. 2012). The goal of this study is to determine whether the major subfamilies of SNF2 and the four components that make up the SWI/SNF core complex are conserved in *Chondrus crispus*, and whether they are expressed in a manner that is consistent with the formation of a classic SWI/SNF complex.

Methods

My first task was to try to identify which SNF2 proteins are conserved in the *Chondrus crispus* genome. This was completed using the National Center for Biotechnology Information (NCBI) website (<http://www.ncbi.nlm.nih.gov/>) where the genome for *Chondrus* has recently been deposited. I used the NCBI's database BLAST function to search for my targeted protein sequences. I searched for all the subfamilies of the larger SNF2 superfamily as well other proteins from the SWI/SNF, and INO80 chromatin remodeling complexes. Within the SWI/SNF complex, I explored the *Chondrus* genome for the four core proteins that form it (Snf2, Swi3, Swp73, and Snf5). In order to find my desired protein sequences I completed multiple BLASTp searches of the *Chondrus* genome using known sequences from model organisms where SWI/SNF functions have been characterized. BLAST stands for the Basic Local Alignment Search Tool and finds regions of local similarity between sequences. This tool is used for the comparison of protein and nucleotide sequences to sequence databases and calculates statistical significance of matches. *Saccharomyces cerevisiae*, *Homo sapiens*, and *Drosophila melanogaster* were used to search for SNF2 sequences since they have already been identified on NCBI. I then kept the putative homologs with an E-value less than to the -20 to ensure that I had an accurate match to the desired sequence. Once I had the sequence I completed a reciprocal blast. This is when I used the sequence I found in *Chondrus* as a query in a complete BLASTp search against *Saccharomyces cerevisiae*, *Homo sapiens*, and *Drosophila melanogaster* to ensure that I had the desired protein sequence from *Chondrus*. Each sequence was saved as an individual text file.

Once I obtained all the sequences from the initial search of the *Chondrus* database for SNF2 superfamily sequences, I compiled a text file that included each sequence that had an E-value of less than to the -20. I then received a file from Dr. Stiller that had known sequences from each SNF2 subfamily for many different model organisms. I compiled the sequences I found from *Chondrus* with the sequences that Dr. Stiller gave me. MEGA6 a computer program an interactive alignment tool was downloaded and used. MEGA6 software was found at <http://www.megasoftware.net/>. This software allowed me to see how my genes and sequences align compared to the known sequences. Once all the sequences were loaded into MEGA6 I carefully went through and removed sequences that did not align well in the conserved SNF2 domains, and trimmed regions of the alignment with poorly aligned gaps or missing data. After editing the alignment I would realign it and repeat the process until I got an unambiguous alignment across all organisms. This file was then saved into a FASTA file. Once a final alignment was achieved the file was selected and a maximum-likelihood tree was constructed using MEGA. Bootstrap analyses were used to assess statistical support for the nodes on the tree with 100 replications to ensure accuracy. After completion of the tree I had to categorize the subfamilies identified by each branch of the tree (**Figure 2**). Classification was completed using a previously published tree as a guide for family determination (Stiller et al. 2012). The next step was to label the tree based on subfamily classification

My next step was to try to determine whether the SWI/SNF complex was formed in *Chondrus*. In order to do this, the four major groups of proteins that make up the SWI/SNF multi-subunit (Snf2/swi2, Swi3, Swp73, and Snf5) were targeted. Using known sequences from *Saccharomyces cerevisiae*, *Homo sapiens*,

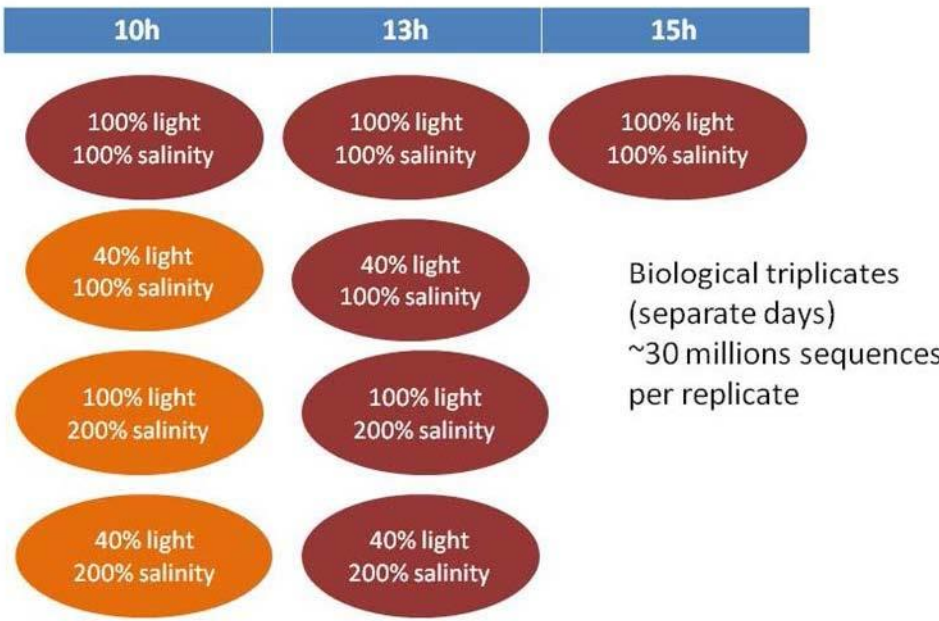
and *Drosophila melanogaster*, BLASTp searches of the *Chondrus* genome were completed. Sequence matches that received an E-value less than 10^{-20} were kept and a reciprocal blast of each was completed to ensure accuracy. Once all the sequences were obtained they were converted into a text file. The same process was completed for the proteins involved in the INO80 complex. The sequences for the SWI/SNF complex and the INO80 complex were compiled into a text file. This file was sent to a collaborator in France by the name of Jonas Collén to search the *Chondrus* project database for gene expression data. Collén returned expression data for the four major protein sequences that make up the SWI/SNF complex as well as the sequences of the INO80 complex.

The expression data were obtained from multiple environments with various levels of salinity, light exposure, and at different time intervals as shown in **Figure 1**. The first data was taken at 10 hours at a 100% light and at 100% salinity. The next condition was at 13 hours at a 100% light and 100% salinity. The third was at 15 hours at a 100% light and 100% salinity. These three conditions were replicated three times to ensure accuracy and to look for correspondence. For the next conditions the amount of light exposure and the levels of salinity were changed. The fourth environment from which expression data was taken was at 13 hours under 40% light and a 100% salinity. The fifth was at 13 hours with a 100% light and 200% salinity. The sixth and final environment was at 13 hours, 40% light and 200% salinity. Upon receiving the expression data I averaged the values from the three replicates at each condition (shown in **Table 3**). These values were analyzed to see whether the data were consistent with formation of a SWI/SNF complex in *Chondrus* and to see if there was any correlation between the expression of the SWI/SNF and INO80 complex.

Figure 2

Figure 2. Maximum-likelihood tree of SNF2 sequences from *Porphyra*, *Chondrus*, *Cyanidioschyzon* and model organisms in which gene/protein functions have been characterized. Sequences from red algae are shown in red. Pum followed by numbers represents *Porphyra*. CM stands for *Cyanidioschyzon* and snf2 followed by numbers represents *Chondrus*. Hsa-*Homo sapiens*, Ath-*Arabidopsis thaliana*, Sce-*Saccharomyces cerevisiae*, Mmu-*Mus musculus*, Dme-*Drosophila*

Stress in the natural environment



Doctoral thesis
Nathalie Kowalczyk

Figure 3
Figure 3. This displays the conditions from which expression data were taken. Samples are indicated by the brown circles. Experiments were started at 10h. Letters correspond to the letters above the columns inTable 1.

	A	B	C	D	E	F
snf2	0.81	-1.46	0.03	0.56	-1.34	-0.62
swi3	0.17	-0.51	0.26	0.51	-0.62	-0.24
swp73	Not found					
snf5	0	-0.54	0.54	0.21	-0.61	-0.06
RVB	0.61	-1.59	0.24	0.63	-1.57	-0.72
RVB	0.29	-0.82	0.25	0.44	-0.57	-0.16
INO80	0.75	-1.48	-0.32	0.57	-1.02	-0.33

Table 1

Table 1. Column A, B, and C show the mean relative expression values versus the overall average of three replicates under 100% light and 100% salinity at A (10hr), B (13hr), and C (15h). Conditions of the experiment are found below in Figure 2. At a value of 0 the genes are expressed at average/normal levels of expression. Positive values show the genes are being expressed above average. Negative values indicate the genes are under expressed at those given conditions.

Results

The maximum-likelihood phylogenetic tree, shows where *Chondrus crispus* homologs align in comparison to other homologs from known organisms (Figure 2). As displayed on the tree, the homologs of *Chondrus* typically are found to be most closely related to other red algal sequences. When other red algal sequences are not found present in the subfamily, we can see which other organism the sequence is closely related too. It was also found that *Chondrus* had conserved a homolog from each of the major subfamilies of the larger SNF2 superfamily. A few homologs were found that did not obviously group with known SNF2 subfamilies. Their functions are not known and possibly could be uncharacterized paralogs or sequences whose functions perform similarly to those of other SNF2 proteins.

Sequences for each of the four major proteins that form the core SWI/SNF complex, as well as sequences that form the INO80 complex were found to be conserved in *Chondrus*. Experimental expression data revealed that Swp73 was not expressed. However, expression data for Snf2, Snf5, and Swi3 appeared to be correlated with the genes that form the INO80 complex. The three other SWI/SNF proteins (other than Swp73) show comparable patterns of increased and decreased expression with each other across the samples. Fifteen out of the eighteen replicates were found to have comparable above or below average expression data between the protein sequences that form the INO80 and SWI/SNF complex.

Discussion

After searching the complete genome of *Chondrus*, members of most SNF2 helicases sub-families were found to be present; however the absence of expression of a key gene, Swp73, leads to doubt that the classic SWI/SNF complex is formed. Upon analyzing the expression data, there seems to suggest that there may be correlation between the genes that form the SWI/SNF complex and those that form the INO80 complex. These findings are very interesting because the complexes may be formed in *Chondrus*. If this is indeed the case then it is very likely that the complexes evolved independently in red algae. Even though SWP73 is not expressed with the other components of the SWI/SNF complex, the expression data suggest that the three other genes that form the complex appear to have correlated expression. This correlation suggests that they could be expressed in a chromatin remodeling complex together. More research is needed to replicate the data and findings presented by this study. These findings could give some insight to the evolution of *Chondrus* and red algae and to why red algae failed to develop differential tissue and more complex cellular complexes that are typically found in multicellular eukaryotes.

Reference Page

- Archacki, R., Sarnowski, T. J., Halibart-Puzio, J., Brzeska, K., Buszewicz, D., Prymakowska-Bosak, M., Koncz, C. & Jerzmanowski, A. 2009. Genetic analysis of functional redundancy of BRM ATPase and ATSWI3C subunits of Arabidopsis SWI/SNF chromatin remodelling complexes. *Planta* 229:1281-92.
- Collén, J., Porcel, B., Carré, W., Ball, S. G., Chaparro, C., Tonon, T., et al. (2013). Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the archaeplastida. *Proceedings of the National Academy of Sciences USA*, 110(13), 5247-5252.
- Eisen, J. A., Sweder, K. S. & Hanawalt, P. C. 1995. Evolution of the SNF2 family of proteins subfamilies with distinct sequences and functions. *Nucl. Acids Res.* 23:2715–23.
- Flaus, A., Martin, D. M. A., Barton, G. J. & Owen-Hughes, T. 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucl. Acids Res.* 34:2887–905.
- Introduction to rhodophyta*. (1999). Retrieved 03/22, 2014, from <http://www.ucmp.berkeley.edu/protista/rhodophyta.html>
- Jerzmanowski, A. 2007. SWI/SNF chromatin remodeling and linker histones in plants. *Biochim. Biophys. Acta - Gene Structure and Expression* 1769:330–45.
- Pazin, M. J., & Kadonaga, J. T. (1997). SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell*, 88(6), 737-740.
- Rayment, W., & Pizzola, P. *Chondrus crispus*. *Carrageen*. Retrieved 04/16, 2014, from <http://www.marlin.ac.uk/speciesfullreview.php?speciesID=2971>

Stiller, J. W., Perry, J., Rymarquis, L. A., Accerbi, M., Green, P. J., Prochnik, S., et al. 2012. Major developmental regulators and their expression in two closely related Species of *Porphyra* (Rhodophyta). *Journal of Phycology*, 48(4), 883-896.

Winston, F. & Carlson, M. 1992. Yeast SNF SWI transcriptional activators and the SPT SIN chromatin connection. *Trends Genet.* 8:387–91.